

HIV evolution: CTL escape mutation and reversion after transmission

A J Leslie^{1,12}, K J Pfafferoth^{1,12}, P Chetty², R Draenert³, M M Addo³, M Feeney³, Y Tang³, E C Holmes⁴, T Allen⁵, J G Prado⁵, M Altfeld⁵, C Brander⁵, C Dixon¹, D Ramduth², P Jeena², S A Thomas², A St John⁶, T A Roach⁶, B Kupfer⁷, G Luzzi⁸, A Edwards⁹, G Taylor¹⁰, H Lyall¹⁰, G Tudor-Williams¹⁰, V Novelli¹¹, J Martinez-Picado⁵, P Kiepiela², B D Walker³ & P J R Goulder^{1,3}

Within-patient HIV evolution reflects the strong selection pressure driving viral escape from cytotoxic T-lymphocyte (CTL) recognition. Whether this intrapatient accumulation of escape mutations translates into HIV evolution at the population level has not been evaluated. We studied over 300 patients drawn from the B- and C-clade epidemics, focusing on human leukocyte antigen (HLA) alleles *HLA-B57* and *HLA-B5801*, which are associated with long-term HIV control and are therefore likely to exert strong selection pressure on the virus. The CTL response dominating acute infection in *HLA-B57/5801*-positive subjects drove positive selection of an escape mutation that reverted to wild-type after transmission to *HLA-B57/5801*-negative individuals. A second escape mutation within the epitope, by contrast, was maintained after transmission. These data show that the process of accumulation of escape mutations within HIV is not inevitable. Complex epitope- and residue-specific selection forces, including CTL-mediated positive selection pressure and virus-mediated purifying selection, operate in tandem to shape HIV evolution at the population level.

Evasion of the host CTL response through mutation of key epitopes is a major challenge to natural or vaccine-induced immune control of HIV^{1–4}. This phenomenon of CTL escape mutation in immunodeficiency virus infection has been well documented^{1–10}. Recent studies suggest that, through escape mutation, CTL has a major role in driving evolution of HIV^{5,9,10}. However, the extent to which this actually occurs remains to be determined. It is clear that CTL escape mutations are common, and are selected by CTL responses restricted by a wide array of different HLA molecules⁹. However, the question of whether these mutations are transmissible and then stable in the absence of the HLA allele that drove their selection has yet to be addressed. These two parameters need to be established to quantify the role of the CTL response in driving HIV evolution at a population level. It is also of vital importance in terms of vaccine design to determine which immune responses will remain relevant and which will be irrevocably lost through escape mutations.

To address this question, we sought examples of *HLA-B57*- and *HLA-B5801*-restricted CTL escape, and observed the outcome of transmission of these mutations to *HLA-B57/5801*-negative individuals. These HLA alleles were chosen because *HLA-B57* and the

closely related *HLA-B5801* are associated with effective control of HIV^{11–15} (P.J.R.G. *et al.*, unpublished data). We therefore hypothesized that the *HLA-B57/5801*-restricted CTL response represents a strong selection force that would be associated with escape mutation. We focused here on the Gag polyprotein, as this is a highly immunogenic region of HIV-1 (refs. 16–18).

TSTLQEQIAW variation is associated with *HLA-B57/5801*

Gag sequences from 311 subjects with chronic HIV infection were analyzed. Variation in the epitope TSTLQEQIAW (designated TW10; Gag HXB2 residues 240–249) was observed in association with expression of *HLA-B57* or *HLA-B5801* (Table 1a). TW10 dominates the CTL response in acute HIV infection in *HLA-B57* individuals¹⁹, and is presented by the closely related *HLA-B5801* allele¹⁴. The dominant change in TW10 is a substitution of Thr by Asn at residue 242 (T242N). This mutation occurs in C-clade infections, arising in 84% and 63% of *HLA-B57*-positive and *HLA-B5801*-positive subjects, respectively, and in 0% of *HLA-B57/5801*-negative subjects ($P < 0.0001$ in each case). Data from B-clade-infected subjects reveal a similar picture (Table 1a). Again, the T242N mutation is highly prevalent in *HLA-B57*-posi-

¹Department of Pediatrics, Fulfield Department of Medicine, Peter Medawar Building for Pathogen Research, University of Oxford, Oxford OX1 35Y, UK. ²HIV Pathogenesis Programme, The Doris Duke Medical Research Institute, University of Natal, Congella 4013, Durban, South Africa. ³Partners AIDS Research Center and Harvard Medical School, Boston, Massachusetts 02129, USA. ⁴Department of Zoology, University of Oxford, Oxford OX1 35Y, UK. ⁵IRS/Caixa Foundation, University Hospital 'Germanes Trias i Pujol', 08916 Barcelona, Spain. ⁶Queen Elizabeth II Hospital, Bridgetown, Barbados. ⁷University of Bonn, 53105 Bonn, Germany. ⁸Department of Genitourinary Medicine, High Wycombe General Hospital, HP11 2TT, UK. ⁹The Harrison Clinic, Radcliffe Infirmary Hospital, Oxford, OX2 6HE, UK. ¹⁰Imperial College School of Medicine, St Mary's Hospital, London W2 1NY, UK. ¹¹Great Ormond Street Hospital for Sick Children, London WC1N 3JH, UK. ¹²These authors contributed equally to this work. Correspondence should be addressed to P.J.R.G. (philip.goulder@ndm.ox.ac.uk).

Published online 8 February 2004; doi:10.1038/nm992

tive and *HLA-B5801*-positive subjects (79% and 93%, respectively), and is not found in any *B57/5801*-negative individuals ($P < 0.0001$ in each case).

An interclade difference exists within the TW10 epitope at residue 248, where the consensus is Gly (G) in B-clade strains and Ala (A) in C-clade strains. In C-clade-infected subjects, variation at this position is weakly associated with *HLA-B57* only ($P = 0.03$). By contrast, there is a high degree of variation at this position in B-clade-infected subjects, predominantly involving the G248A change, again strongly associated with *HLA-B57* ($P < 0.0001$) but not *HLA-B5801*. Unlike T242N, G248X is found in substantial numbers in *HLA-B57/5801*-negative individuals (Table 1a). Together, these results indicate that TW10 may be under immune selection pressure that results predominantly in a T242N substitution in both B-clade and C-clade infection, as well as an additional G248X substitution, particularly in B-clade infections.

TW10 variation arises after transmission

To show that the TW10 variants arise after transmission to *HLA-B57/5801*-positive subjects, we sought examples of vertical and horizontal transmission in which *HLA-B57/5801*-negative donors, infected with wild-type virus, transmit to subjects expressing *HLA-B57* or *HLA-B5801*. Six such mother-child transmission pairs were identified, with viral sequences obtained from all six mothers encoding wild-type TW10 and from five of the six *HLA-B57/5801*-positive children carrying the T242N and/or G248A mutations (Table 1b). In the two adult pairs in which the direction of horizontal transmission was unambiguous, only the *HLA-B57/5801*-positive subjects possessed the T242N mutation. Likewise, in three additional adult pairs in which the direction of horizontal transmission was ambiguous, only virus isolated from the *HLA-B57/5801*-positive subjects possessed the T242N mutation (Table 1c). In this last group, the data could indicate either transmission of wild-type virus to the *B57/5801*-positive individual followed by escape, or transmission of the T242N mutant virus followed by reversion to wild type in the *HLA-B57/5801*-negative recipient. The authenticity of all of these transmission pairs was confirmed by phylogenetic analysis (Supplementary Fig. 1 online).

T242N arises as a result of positive selection

To determine whether TW10 is under positive selection from

Table 1 Variation in TW10

a	HLA-B57			HLA-B5801			Other		
	TSYTLQRIQIWN	n	%	TSYTLQRIQIWN	n	%	TSYTLQRIQIWN	n	%
C-clade (n = 183)	-N-----	17	53	-N-----	24	56	-N-----	90	83
	-N-----T	5	16	-N-----T	8	19	-N-----T	7	6
	-N-----A	3	9	-N-----A	3	7	-N-----A	3	3
	-N-----M	2	6	-N-----M	2	5	-N-----M	3	3
	-N-----V	1	3	-N-----V	1	2	-N-----V	1	1
	-N-----Q	1	3	-N-----Q	1	2	-N-----Q	1	1
	-N-----D	1	3	-N-----D	1	2	-N-----D	1	1
	-N-----E	1	3	-N-----E	1	2	-N-----E	1	1
	-N-----G	1	3	-N-----G	1	2	-N-----G	1	1
	-N-----T	1	3	-N-----T	1	2	-N-----T	1	1
Total variation 100% (P < 0.0001)			Total variation 81% (P < 0.0001)			Total variation 17%			
T242N 84% (P < 0.0001)			T242N 94% (P < 0.0001)			T242N 9%			
A248X 25% (P = 0.03)			A248X 9% (NS)			A248X 10%			
B-clade (n = 128)	TSYTLQRIQIWN	n	%	TSYTLQRIQIWN	n	%	TSYTLQRIQIWN	n	%
	-N-----A	14	41	-N-----A	6	53	-N-----A	55	70
	-N-----G	9	26	-N-----G	5	2	-N-----G	4	5
	-N-----T	4	12	-N-----T	1	7	-N-----T	2	2
	-N-----D	2	6	-N-----D	1	7	-N-----D	1	1
	-N-----E	1	3	-N-----E	1	7	-N-----E	1	1
	-N-----Q	1	3	-N-----Q	1	7	-N-----Q	1	1
	-N-----V	1	3	-N-----V	1	7	-N-----V	1	1
	-N-----M	1	3	-N-----M	1	7	-N-----M	1	1
	-N-----T	1	3	-N-----T	1	7	-N-----T	1	1
Total variation 97% (P < 0.0001)			Total variation 100% (P < 0.0001)			Total variation 30%			
T242N 79% (P < 0.0001)			T242N 95% (P < 0.0001)			T242N 9%			
G248X 71% (P < 0.0001)			G248X 47% (NS)			G248X 27%			

b	Transmission pair	HLA type	Route	Time point	TSYTLQRIQIWN
	P1750214-Mother 900918KF-Child	B*24:44 B*57:02	Vertical	2-3 years pp* 2 years pp 3 years pp	----- -----A -----A
	B56605-Mother 000401CM-Child	B*27:01 B*27:01	Vertical	2 years pp 2 years pp	----- -----
	DF680505-Mother 951111MP-Child	B*27:03 B*58:01/53	Vertical	6 years pp	----- -----
	Don-17 Don-23	B*44:49 B*57:08	Horizontal	NK*	----- -----
	Bonn-F Bonn-G	B*7: B*57:27	Blood-borne	11 years*	----- -----
	TSYTLQRIQIWN				
	J120-Mother J120-Child	B*42:81 B*57:42	Vertical	6 years pp	----- -----T
	J143-Mother J143-Child	B*44: B*58:01/44	Vertical	4 years pp	----- -----
	J3-Mother J3-Child	B*53:71 B*58:01/71	Vertical	NK	----- -----
	c	043M-Wife 043D-Husband	B*35:49 B*58:01/27	Horizontal	DNA
068M-Wife 068D-Husband		B*27:53 B*58:01/72	Horizontal	DNA DNA DNA	----- -----N -----N
600E-Wife 600F-Husband		B*27:44 B*58:01/53	Horizontal	DNA	----- -----N
Population averaging					
7/7 5/5 3/4 1/4					
Population averaging					
7/7 5/5 3/4 1/4					

(a) All sequences are from HIV antibody-positive individuals, and represent population sequences from proviral DNA. NS, not significant. (b) Transmission of wild-type TW10 to *B57/5801*-positive subjects in unambiguous direction. J120, J143 and J3 mother-and-child pairs were infected with C-clade HIV; remaining transmission pairs were infected with B-clade HIV. Don-17 and Don-23 are a previously reported horizontal transmission pair⁴⁴, in which patient history indicates that Don-17 infected Don-23. Bonn-F and Bonn-G, also previously reported⁴⁵, are hemophiliacs infected with the same batch of contaminated factor VIII. An sequence data shown represent population sequences from proviral DNA. *Postpartum. †Time after transmission not known. ‡Time after transmission. B57 and B5801 are shown in bold. (c) Transmission of HIV within husband-wife pairs in which the direction of transmission was unknown. All subjects were infected with B-clade virus; time of transmission is unknown.

HLA-B57/5801-restricted CTLs, we examined selection pressures using a maximum-likelihood method²⁰. This method revealed nine sites under positive selection when *B57/5801*-positive and non-*B57/5801* subjects were grouped together (Supplementary Table 1 online), including residues 242 (mean ratio of nonsynonymous to synonymous nucleotide substitutions (d_N/d_S) = 3.216)

and 146 (mean $d_{50}/d_{50} = 3.215$). Both results were highly significant, with Bayesian posterior probabilities of 1.000 and 0.999, respectively. When *HLA-B*57:801*-positive subjects were excluded from the analysis, only residues 242 and 146 were no longer identified as being under positive selection, suggesting that selection at these sites is operating through the *HLA-B*57:801* alleles. The same analysis conducted using the B-clade sequence data similarly identified residues 242 and 146 as being under positive selection (mean $d_{50}/d_{50} = 2.34$ and 2.23, respectively).

TW10 variants are *B57/801* CTL escape mutations

The optimal epitope TW10 (TSTLQEQIGW) was previously defined in B-clade-infected subjects expressing *HLA-B*57* or *B*801* (ref. 14). TW10 (TSTLQEQIGW) was confirmed as the optimal epitope in C-clade infection (Fig. 1a). Recognition of the commonly occurring variants at residues 242 and 248 was evaluated in *B57*- and *B*801*-positive subjects with B- or C-clade infection, respectively (Fig. 1b–e). The pattern of recognition was similar in each case, with reduced recognition equivalent to 1–2 logs peptide concentration of variants with single amino-acid substitutions, and complete abrogation of a response to peptides bearing double mutations. Thus, these *HLA-B*57*- and *B*801*-associated mutations within TW10, which dominate the viral population in chronically infected individuals expressing these alleles, represent escape mutations.

T242N 'reverts' in the absence of *B57/801*

We noted above that T242N is seen in 0 of 187 *HLA-B*57/801*-negative subjects with chronic HIV infection. The most likely explanation for this is that T242N mutant viruses are not transmitted, or that the T242N mutant is outcompeted by wild-type virus ('reverts') after transmission to *B57/801*-negative individuals. To determine whether T242N is transmitted, instances of transmission from *HLA-B*57/801*-positive subjects to individuals lacking these alleles were sought. In one such case of intrapartum mother-to-child transmission (Fig. 2a), the B-clade-infected mother (SMH-05M; *HLA-B*57*-positive) carried both the T242N and G248A mutations (46 of 46 clones) at 5 months, 7 months and 8 years postpartum, indicating that both mutations are stable in a

B57/801-positive individual for this duration. In the child (SMH-05C; *HLA-B*7* homozygous) at 2 months postpartum, all clones possessed the T242N and G248A mutations, confirming that both mutations were transmitted. By 5–7 months of age, the frequency of the T242N mutation declined to ~50%, and by 5 years none of the 16 clones isolated from plasma RNA possessed the T242N mutation. (Table 2 and Fig. 2a). Although 3 of 17 clones isolated from proviral DNA at 8 years of age encoded T242N, phylogenetic analysis confirmed that these are closely related to the transmitted virus, and thus represent archival sequences (Fig. 2b). In contrast to T242N, the G248A mutation was present in all SMH-05C clones from all time points, suggesting that G248A is stable (does not revert) in the absence of the selecting *HLA* allele.

A second mother-to-child transmission pair, 997-Mother (*HLA-B*18/57*) and 997-Child (*HLA-B*18/42*), was identified (Table 2 and Fig. 2c). This pair was C-clade infected. T242N was present in the mother antenatally and at all time points postpartum, as expected. In the child, who was infected 3–9 months postpartum, 0 of 22 clones sequenced from RNA had the T242N mutation at the 9-month time point. These data again support the hypothesis that the T242N mutation reverts in the absence of *HLA-B*57/801*.

Analysis of 187 chronically infected *HLA-B*57/801*-negative individuals revealed none with virus expressing the T242N mutation. In contrast, virus expressing the T242N mutation was isolated from 4 of 19 *HLA-B*57/801*-negative subjects with acute HIV infection (Table 2). This further demonstrates that T242N is transmitted horizontally, and is in itself evidence that T242N reverts at some point between acute and chronic infection in *HLA-B*57/801*-negative individuals in whom the transmitted virus carries the T242N mutation.

One of these acutely infected individuals, AC-33, was studied longitudinally, revealing that T242N also reverts over time in this situation (Table 2 and Fig. 2d). This subject was placed on antiretroviral therapy (ART) immediately after diagnosis of HIV infection. During ART, viral replication was suppressed to undetectable levels (<50 copies/ml). One might thus anticipate that T242N reversion would take longer in this setting than in patient SMH-05C, who did not receive ART until 4 years after infection, or in patient 997-C, who never received ART.

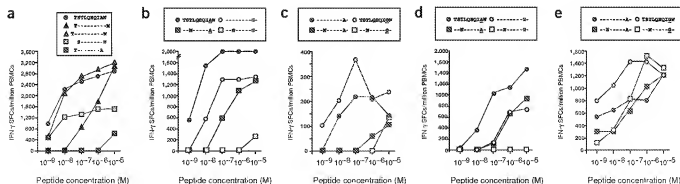


Figure 1 TW10 variants are escape mutations. IFN- γ production was recorded as spot-forming cells (SFCs) per million PMBCs. Results shown are representative of assays using PBMCs from seven subjects expressing either *HLA-B*57* or *HLA-B*801*. (a) Recognition of TW10 and suboptimal 9-mer and 11-mer overlapping peptides using PBMCs from C-clade-infected donor A-005-M (*HLA-B*57/7*). (b–d) TW10 variant recognition. (b) *HLA-B*57*-positive, C-clade-infected subject A-005-M (c), *HLA-B*57*-positive, B-clade-infected subject SWS (*HLA-B*57*; d), *HLA-B*801*-positive, C-clade-infected subject PS-032-M (*B*801*); e) and *HLA-B*801*-positive, B-clade-infected subject, 90-92851-R1 (*B*801*). Sequence of wild-type TW10 is shown in full in each key.



Reversion revealed by B57/5801 'footprints'

To further substantiate the hypothesis that T242N undergoes reversion in the absence of HLA-B57/5801, we next sought T242N-linked mutations that are stable in the absence of B57/5801. The term 'footprints'²¹ was previously coined to describe sequence polymorphisms associated with particular HLA alleles. Thus, the presence of characteristic polymorphisms can indicate the impact of specific HLA alleles on HIV. In HLA-B57/5801-negative individuals, the association of such footprints of the T242N mutation with variation

in TW10, which we have shown is strongly associated with HLA-B57/5801, would support the hypothesis that the ancestral virus in these subjects originated in HLA-B57/5801-positive individuals (Fig. 3). Two sites were identified at which variation from the consensus sequence was associated with T242N in the HLA-B57/5801-positive group: H219X (X = Q or R; $P < 0.0001$) and A146X (X = P, T, S, V, N or H; $P = 0.0008$). In the case of A146R, this polymorphism represents a processing escape mutation (R.D. *et al.*, unpublished data). In the case of H219X, no HLA-B57 or B5801-restricted

epitope has been identified for this region^{14,18,22}, and there is no potential B57 or B5801-restricted epitope that would satisfy the defined HLA-B57/5801-binding motif⁶³.

The H219X variant appears subsequent to the development of the T242N mutation. In B57/5801-positive subjects, H219X is seen only in the presence of T242N (32 of 32 instances; Table 3a), compared with 64 cases where T242N is found in the absence of H219X and 28 cases where neither variant is found ($P > 0.0001$). In addition, longitudinal sequencing of virus in SMH-05M demonstrated the acquisition of this H219X variant after fixation of the T242N mutation (Table 3b). The hypothesis that the H219 variant may be maintained after transmission, whereas the T242N variant reverts, is also supported by analysis of two husband-wife pairs (6007/6008 and 068D/068M; Table 3b). In both cases, the T242N mutation present in the putative donors (6007 and 068D; both HLA-B5801-positive) was linked with the H219Q variant, whereas in the spouses (6008 and 068M; both HLA-B5801-negative), only the footprints of these variants remain.

We next examined the B57/5801-negative subjects to test the hypothesis that the T242N footprints H219X and A146X are associated with TW10 mutations other than T242N, as T242N would be expected to revert in the absence of HLA-B57/5801. We found that both H219X and A146X were associated with variation in TW10 ($P = 0.0009$ and $P = 0.0002$, respectively), in spite of the absence of T242N from these subjects. These data, which show an association between the two identified T242N footprints, H219X and A146X, and TW10 variation within HLA-B57/5801-negative subjects, support the hypothesis that the ancestral virus in these individuals originated from HLA-B57/5801-positive subjects in which the T242N variant was present. Taken together, these data provide compelling evidence that the T242N mutation is transmitted to, and subsequently reverts in, B57/5801-negative subjects.

Table 2 Fate of TW10 variants after transmission

Subject	HLA type	Time point	Genetic material	TSTLQKQIGW	Number of clones
SMH-05-Mother	B57/7	5 months	DNA	--N-----A-	16/16
		7 months	RNA	--N-----A-	17/17
		8 years	DNA	--N-----A-	13/13
			RNA	--N-----A-	Population sequencing
SMH-05-Child	B7/-	2 months	DNA	--N-----A-	14/14
		5 months	DNA	--N-----A- -----A-	12/21 9/21
		7 months			
			RNA	--N-----A- -----A-	6/12 6/12
		5 years	RNA	-----A-	16/16
		8 years	DNA	--N-----A- -----A-	3/17 14/17
				TSTLQKQIGW	
997-Mother	B57/18	-8 weeks	DNA	--N-----	Population sequencing
		1 week	DNA	--N-----	Population sequencing
		6 weeks	DNA	--N-----	Population sequencing
		3 months	DNA	--N----- RNA	Population sequencing 15/15
		9 months	DNA	--N-----	Population sequencing
997-Child	B42/18	9 months	RNA	-----M--	22/22
				TSTLQKQIGW	
AC-04	B18/44	Pre-sero	DNA	--N-----A-	Population sequencing
AC-27	B8/35			--N-----	
AC-14	B8/62			--N-----A-	
AC-33	B7/44	Day 0	RNA	--N-----	6/7
		Pre-sero		--N-----M--	1/7
		Day 43	DNA	--N-----	8/8
		Day 818	RNA	--N----- -----	2/7 5/7
		Day 983	RNA	--N----- -----	3/14 11/14
		Day 1541	RNA	-----	4/4

SMH-05-Mother/SMH-05-Child and 997-Mother/997-Child are vertical transmission pairs in which HLA-B57-positive mothers (infected with B-clade and C-clade virus, respectively) transmitted virus to their HLA-B57/5801-negative children. DNA, sequence obtained from proviral DNA; RNA, sequence obtained from viral RNA. All time points are postpartum, apart from one time point (~8 weeks) at which the 997-Mother sample was antenatal. AC-04, AC-27, AC-14 and AC-33 are HLA-B57/5801-negative adults acutely infected with B-clade virus possessing T242N. All were treated with ART in acute infection. Subject AC-33 was followed longitudinally, with time points given as days since seroconversion. Pre-sero, before seroconversion. B57 and B5801 are shown in bold.



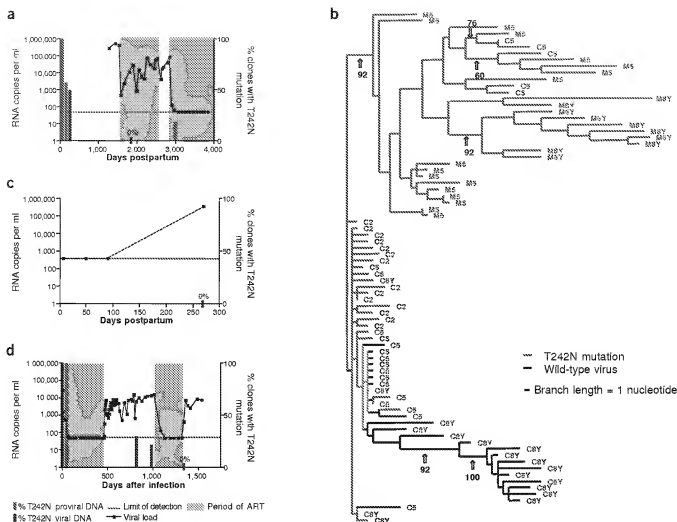


Figure 2 Reversion of TW10 variants after transmission to HLA-B*57:5801-negative subjects. **(a, c, d)** Viral loads of three subjects, SMH-05C **(a)**, 997C **(c)** and AC-33 **(d)**, showing intrapartum mother-to-child, postpartum mother-to-child and horizontal transmission, respectively. **(b)** Maximum-likelihood phylogenetic tree of SMH-05-Mother and SMH-05-Child proviral DNA clones. Blue, SMH-05-Child clones from 2 months (C2), 5 months (C5) and 8 years (C8Y); pink, SMH-05-Mother clones from 5 months (M5) and 8 years (M8Y); T242N mutation is mapped onto the tree. Lineages in which ancestral state is equivocal are highlighted in gray. Yellow arrows indicate major sublineages supported by high bootstrap values.

DISCUSSION

These data provide the first clear evidence of reversion after transmission of an HIV-1 CTL_E epitope in which escape has occurred in the donor. We show that T242N is an escape mutation in the TW10 epitope that is selected in the majority of individuals possessing either HLA-B*57 or HLA-B*5801. B*57:5801-positive individuals can transmit the T242N mutant, and such transmission has occurred in numerous instances. The T242N mutation was found in 0 of 187 of HLA-B*57:5801-negative individuals with chronic HIV infection, compared with 4 of 19 individuals in the acute phase of infection ($P = 0.00005$), providing strong evidence that T242N is transmitted and, in the absence of HLA-B*57:5801, reverts to wild type after acute infection. We present longitudinal data from three HLA-B*57:5801-negative subjects to demonstrate that the transmitted T242N variant is replaced by wild-type virus over time. Further evidence of T242N transmission, followed by reversion, derives from the analysis of genetic footprints that are linked with T242N in HLA-B*57:5801-positive subjects, and that can be identi-

fied in B*57:5801-negative subjects in whom T242N is no longer present.

A previous study examining escape mutations in the HLA-B*27-restricted epitope KK10, a second epitope associated with long-term control of HIV infection, indicated that the occurrence of reversion after transmission in this instance is rare¹⁰. Intrapatient reversion after progression to AIDS is also rare, having been suggested by only a single clone out of >150 clones analyzed from that study and others^{10,24}. Thus, the B*27-KK10 escape variants generally remain stable in the absence of the evolutionary pressure that originally selected them^{10,24}. This contrasts with T242N, which consistently reverts after transmission to a B*57:5801-negative individual. This distinction highlights the importance of determining the fate of CTL escape mutations. If they are transmissible and stable after transmission, they will spread through the population and the epitope will eventually become extinct. Conversely, if the escape mutant reverts, the epitope will remain useful for CTL-based vaccine design.

The likely explanation for T242N reversion is that there is an associated fitness cost. Drug-resistance mutations present an obvious analogy, as they typically do not persist in untreated patients, despite conferring greater fitness in the presence of ART, because they impair the replication of wild-type virus²⁵. Thus, ART-resistant mutants may pay an 'evolutionary penalty' for their survival²⁶. Similarly, the T242N mutation may represent the evolutionary penalty paid by HIV to

escape TW10-specific CTLs. TW10 is the immunodominant CTL response in acute infection¹⁹, so even an attenuated virus would have a selection advantage over one that is recognized and eliminated by TW10-specific CTLs. Recent data indicate that in the rare *HLA-B*57-*

3

HLA-B*57:5801-positive				HLA-B*57:5801-negative			
B-allele				B-allele			
A146X	H219X	TSTLTQRQIGLN	n = 49	A146X	H219X	TSTLTQRQIGLN	n = 79
P	-	-N-----	6	P	-	-----A-	5
P	-	-N-----	4	P	Q	-----T-	2
P	-	-N-----	4	P	-	-----T-	1
P	Q	Q-----A-	4	-	Q	-----S-	1
-	-	-N-----	2	-	Q	-----A-	1
-	-	-N-----T-	2	T	Q	Q-----	1
-	-	-N-----	2	-	-	-----S-	1
P	-	-N-----	2	P	-	-----A-	1
P	-	-N-----	2	-	-	-----A-	1
P	-	-N-----	2	-	Q	-----	4
-	-	-N-----	1	T	-	-----	1
-	Q	Q-----A-	1	V	-	-----	1
-	-	-N-----	5	-	-	-----	45
-	-	-N-----	3	-	-	-----A-	8
-	-	-N-----D-	1	-	-	-----S-	1
-	-	-R-----A-	1	-	-	-----P-	1
-	-	-N-----	1	-	-	-----VA-	1

C-allele				C-allele			
A146X	H219X	TSTLTQRQIGLN	n = 75	A146X	H219X	TSTLTQRQIGLN	n = 108
P	-	-N-----	18	S	Q	-----T-	6
P	Q	-N-----T-	4	P	-	-----T-	1
P	-	-N-----	2	S	Q	Q-----	1
P	Q	Q-----	2	-	Q	-----S-	1
T	-	-N-----	2	-	Q	Q-----Q-	1
-	Q	Q-----T-	1	-	-	-----T-	1
H	Q	Q-----V-	1	P	Q	-----	1
P	-	-N-----V-	1	-	-	-----V-	1
P	R	-N-----	1	-	-	-----G-	1
P	-	-N-----N-	1	-	-	-----	77
P	-	-N-----T-	1	-	-	-----S-	3
P	-	-N-----	1	T	-	-----	3
P	-	-N-----	1	P	-	-----S-	2
P	-	-N-----HT-	1	-	-	-----V-	2
T	S	-N-----	1	-	-	-----VD-	1
-	-	-N-----	16	-	-	-----D-	1
-	-	-P-----	2				
-	-	-Q-----	2				
-	-	-A-----	2				
-	-	-S-----	2				
-	-	-N-----G-	1				
-	-	-N-----V-	1				
P	-	-N-----	1				
-	-	-N-----	7				

b

Subject	Time point	A14EX	H21.9X	T2T1Q2Q2Q2Q	Genotype	Number of cases
SMH-05M (B07007)	5 months pp	P	-	- - - - - A-	DNA	16/16
	8 years pp	P	Q	- - - - - A-	DNA	9/13
026-BMC (B07042)	10 years pp	P	-	- - - - - A-	DNA	4/13
		P	Q	- - - - - A-	DNA	2/5
6007 (B080153)	NK*	-	Q	- - - - - A-	DNA	1/5
6008 (B27044)	NK	-	Q	- - - - - A-	DNA	2/5
068-D (B080172)	NK	-	Q	- - - - - A-	DNA	2/2
	-	-	Q	- - - - - A-	DNA	3/4
	-	-	Q	- - - - - A-	DNA	1/4
068-M (B27053)	NK	-	Q	- - - - - A-	DNA	7/7
	-	-	Q	- - - - - A-	DNA	5/5

(a) H219X and A146X are associated with T242N in the B57/5801-negative group. In the B57/5801-negative group, H219X and A146X are associated with variations in TW10 other than T242N. (b) Patterns of mutation and reversion of TW10-associated variants. Longitudinal analysis of HLA-B57-positive subject SMH-05M shows acquisition of H219Q mutation after fixation of T242N mutation. Clones sequenced from D26-BMC show presence of H219Q only in association with T242N. The two chronically infected husband-wife pairs 6007/6008 and 06BDB/06SM further support the hypothesis that H219Q is maintained, whereas T242N is lost after transmission to a B57/5801-negative donor. ^cPostpartum. ^dTime after transmission not known. B57 and B5801 are shown in bold.

positive individuals in whom the TW10 epitope has not mutated, viremia is successfully suppressed for decades to <50 copies/ml (ref. 27). Once selection pressure from TW10-specific CTLs is removed, as in transmission to a non-B57/5801-expressing individual, one would expect natural selection to return the virus to its fitter, wild-type state. Even minor enhancements of viral fitness can quickly lead to replacement of the less fit variants in a population²⁸.

One might further speculate that the early dominance of the TW10-specific response in acute infection, followed by the emergence of escape variants conferring an attenuated replicative state upon the virus, could be causally linked to the association of *HLA-B57/B5801* with long-term suppression of HIV infection. In our ART-naïve C-clade-infected study group, *HLA-B57/5801*-expressing subjects whose virus encoded T242N had a median viral load of 10,342 HIV RNA copies per ml plasma, still significantly lower than in the *HLA-B57/5801*-negative group (median viral load = 28,125 HIV RNA copies per ml plasma; $P = 0.02$ by Mann-Whitney test). This T242N mutation may therefore contribute to the control of viremia observed in infected persons expressing *HLA-B57/B5801*, despite the loss of CTL recognition through this TW10 specificity.

The precise mechanism by which the T242N mutation might affect viral fitness is unknown. Gag residue 242 falls within the HIV capsid protein, p24, which has crucial roles in viral function²⁹ and is a highly conserved region of the HIV genome³⁰. Single-point mutations in p24 can have a profound effect on viral fitness^{31–34}, and linker-insertion mutations between residues 241–242 and 242–243 destroy HIV's ability to replicate^{32,35,36}. T242N mutant viruses can persist for at least 8 years; however, replicating and evolving (Fig. 2b). Under such circumstances, one might expect HIV to accumulate other mutations that restore viral fitness to some degree, as suggested for the late-developing B27-KK10 escape mutation²⁴. The H219X mutation, universally observed in association with T242N in *B57/5801*-positive subjects, may represent a similar compensatory mutation. If so, however, the fact that 22 *B57/5801*-negative subjects possessed H219X in the absence of T242N suggests that, unlike the compensatory changes associated with KK10 escape, H219X does not fully restore viral fitness. H219 lies within the cytoplasmic A binding loop, which essential to HIV's life cycle^{35–37}. H219 is directly involved in cytoplasmic A binding, and the H219Q mutation results in a 4.8-fold decrease in binding affinity³⁷. It would thus be unlikely for the H219Q mutation to be selected unless it was linked to some separate fitness advantage—in this case, hypothetically, reducing the negative effect of T242N.

In these studies, we have illustrated two contrasting outcomes of CTL escape mutation. In the case of the T242N mutation, which we hypothesize is associated with a fitness cost to the virus, we observe reversion after transmission to *B57/5801*-negative recipients. Conversely, in the case of the G248A mutation, occurring within the same TW10 epitope, we observe maintenance after transmission to recipients lacking *B57/5801*. The inference that this CTL escape mutation arises at an insignificant fitness cost to the virus is supported by a study that examined this particular G248A mutation and found no effect on viral fitness³⁴. Between these two extremes, there is likely to be a broad spectrum of CTL escape mutations whose presence may be associated with greater or lesser degrees of constraint on viral replication. Our findings illustrate that intrapatient evolution of HIV driven by CTL escape does not necessarily translate into evolution of HIV at the population level. The degree to which escape mutations revert or are maintained will determine the extent to which they contribute to the evolution of HIV as it spreads through a population.

In conclusion, we have identified two escape mutations in the immunodominant *HLA-B57/5801*-restricted Gag epitope TW10, one of which undergoes reversion and one of which is stable in the absence of *B57/5801*. This highlights the diverse outcomes of CTL escape mutation, and confirms the need to clarify the fate of escape mutations in order to determine their future relevance to vaccine design. These findings also suggest that the role of CTLs in shaping HIV evolution may not be a simple one, and that intrapatient viral evolution does not necessarily translate to evolution of HIV at the population level. Rather, it is the complex interaction between different selection forces—positive selection pressure from CTLs on one hand, and purifying selection in the virus on the other—that determines the role of CTLs in shaping HIV evolution.

METHODS

Subjects. The 128 C-clade-infected subjects analyzed in this study were collected from Durban, South Africa, and consisted predominantly of Zulu and Xhosa women recruited from the Cato Manor antenatal clinic. All subjects were ART naïve. The median viral load was 14,892 HIV RNA copies per ml plasma (range <400–698,000), and the median absolute CD4 count was 449 cells/mm³ (range 101–1,215 cells/mm³). The 97 B-clade-infected subjects were collected from diverse sources encompassing Europe, the Caribbean and North America. Of these subjects, 17% were receiving ART at the time of analysis. The median viral load in those not receiving ART was 24,600 RNA copies per ml plasma (range <400 to >750,000), and the median CD4 count for those not receiving ART was 420 cells/mm³ (range 131–1,280 cells/mm³). This study was approved by institutional review boards, and all subjects gave written informed consent.

Sequencing of proviral DNA and viral RNA. We extracted genomic DNA from peripheral blood mononuclear cell (PBMC) pellets using the Puregene DNA isolation kit (Gentra). We then amplified HIV gag sequences using a nested PCR as described³⁸, using the following gag-specific primers: 5'-CTAGCAGTGGCGCCGCAACA-3' and 5'-ACAGTCTTTGATTTGGTGTC-CTTC-3' for first-round outer PCR, and 5'-TCTCTGCAGCCGAGGACTC-3' and 5'-TTTCCACATTTCCCAACAGCC-3' for second round inner PCR. The PCR product was purified by PEG precipitation and either directly sequenced (referred to in the text as 'population sequencing'), or cloned as previously described³⁸ using a TOPO TA cloning kit (Invitrogen). We recovered viral RNA from plasma samples using the Nucleospin RNA extraction kit (Machery-Nagel). During this step, we added RNA-free DNase (Qiagen) to ensure removal of proviral DNA. We then synthesized a cDNA library from the RNA using the Reverse IT 1st Strand Synthesis kit (ABgene), using random oligonucleotides to prime the reaction. We then amplified and sequenced gag fragments from the cDNA as described above. All sequencing was done with BigDye Terminator v3.0 Ready Reaction mix (Applied Biosystems), using the two inner PCR primers listed above and four additional primers (5'-CTGCACATAGGATAATTTTGAC-3', 5'-GACAC-CAAGGAAGCCCTTAG-3', 5'-CTCCCATCTGGAACAGGTG-3' and 5'-GGAA-CAAATAGCATGGATGAC-3'). Sequences were analyzed on the ABI 3700 DNA analyzer. All residue numbers were taken against the HXB reference sequence.

HLA class I typing. HLA class I typing was performed on extracted genomic DNA by PCR single-strand conformation polymorphism³⁹.

Epitope optimization and recognition of TW10 variant peptides. We confirmed TW10 (TSTLQEQIAW) as the optimal B57-restricted epitope in C-clade infection as previously described⁴¹, using fresh PBMCs from the C-clade infected subject A-005M (*HLA-B57/77*). TSTLQEQIAW was confirmed as the optimal epitope for B-clade infection in the same way, using PBMCs from the B-clade infected subject SWS (*HLA-B57/13*). In both cases, we tested fresh PBMCs for recognition of synthetic TW10 peptides in ELISpot assays as previously described⁴⁰.

Selection analysis and phylogenetic inference. A measure of the selection pressures operating at individual codons in the Gag protein was obtained using a maximum likelihood method⁴⁰, which was recently shown to be a powerful way to determine selection pressures in HIV⁴¹. This method works



by assessing the fit to the data of various models of codon evolution, which differ in how d_3/d_4 varies across the sequence and takes into account the phylogenetic relationships of the sequences in question²⁸. Two models of codon evolution were used: M7, which specifies ten categories of d_3/d_4 in the alignment, none of which may be <1 so that evolution is entirely neutral; and M8, which only differs from M7 in that it incorporates an extra (eleventh) class of codons that can take on any value of d_3/d_4 including those supporting positive selection ($d_3/d_4 > 1$). Because M7 and M8 are nested, they may be compared using a standard likelihood ratio test. Consequently, if model M8 significantly rejects model M7 and includes a class of codons where $d_3/d_4 > 1$, we can conclude that positive selection has acted. Individual positively selected amino acid sites can then be identified using a Bayesian approach. All analyses were undertaken using the CODEML program from the PAML package⁴², with input phylogenetic trees inferred using the maximum likelihood method available in the PAUP* package⁴³. All other phylogenetic trees were also inferred using the maximum likelihood method available in PAUP*, assuming the GTR + Γ + I model of nucleotide substitution (all parameter values are available from the authors on request). To assess the support for individual nodes on the tree, a bootstrap resampling analysis was undertaken, using the same substitution model but with 1,000 replicate neighbor-joining trees.

Note: Supplementary information is available on the Nature Medicine website.

ACKNOWLEDGMENTS

We thank those who made possible the collection of blood samples for this study, in particular the staff at Cato Manor Clinic, Durban (the largest single cohort). We also thank M. Hammond and M. Bunce for HLA typing; C. Hull Jackson, B. Matz, C. Edwards and H. Coovadia for facilitating these studies; M. Maiden and L. Richardson for assisting with the sequencing; and P. Klemmer and K. Phillips for critically reviewing the manuscript. This work was supported by the Wellcome Trust (P.R.G. and A.L.), the Elizabeth Glaser Pediatric AIDS Foundation (P.R.G. and M.F.), the Spanish Health Department (BEI, RD 603/17) to J.M.-P., the National Institutes of Health (AI46995-01A1 and N01-A15442) and the Doris Duke Charitable Foundation. P.R.G. is an Elizabeth Glaser Scientist and B.D.W. is a Doris Duke Distinguished Clinical Science Professor.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 16 November 2003; accepted 8 January 2004

Published online at <http://www.nature.com/naturemedicine/>

- Koenig, S. et al. Transfer of HIV-1-specific cytotoxic T lymphocytes to an AIDS patient leads to selection for mutant HIV variants and subsequent disease progression. *Nat. Med.* **1**, 330–336 (1995).
- Borrow, P. et al. Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. *Nat. Med.* **3**, 205–211 (1997).
- Goulder, P.J. et al. Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. *Nat. Med.* **3**, 212–217 (1997).
- Baruch, D.H. et al. Eventual AIDS vaccine failure in a rhesus monkey by viral escape from cytotoxic T lymphocytes. *Nature* **415**, 335–339 (2002).
- Klemmer, P., Wu, Y. & Phillips, R. HIV: current opinion in escapeology. *Curr. Opin. Microbiol.* **5**, 408–413 (2002).
- Evans, D.T. et al. Virus-specific cytotoxic T-lymphocyte responses select for amino-acid variation in simian immunodeficiency virus Env and Nef. *Nat. Med.* **5**, 1270–1276 (1999).
- Allen, T.M. et al. Tat-specific cytotoxic T lymphocytes select for HIV escape variants during resolution of primary viraemia. *Nature* **407**, 386–390 (2000).
- Phillips, R.E. et al. Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. *Nature* **354**, 453–459 (1991).
- Moore, C.B. et al. Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science* **296**, 1439–1443 (2002).
- Goulder, P.J. et al. Evolution and transmission of stable CTL escape mutations in HIV infection. *Nature* **412**, 334–338 (2001).
- Koslov, R.A. et al. Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. *Nat. Med.* **2**, 405–411 (1996).
- Migueles, S.A. et al. HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. *Proc. Natl. Acad. Sci. USA* **97**, 2709–2714 (2000).
- O'Brien, S.J., Gao, X. & Carrington, M. HLA and AIDS: a cautionary tale. *Trends Mol. Med.* **7**, 379–381 (2001).
- Goulder, P.J. et al. Novel, cross-restricted, conserved, and immunodominant cytotoxic T lymphocyte epitopes in slow progressors in HIV type 1 infection. *AIDS Res. Hum. Retroviruses* **12**, 1691–1698 (1996).
- Tang, J. et al. Favorable and unfavorable HLA class I alleles and haplotypes in Zambians predominantly infected with clade C human immunodeficiency virus type 1. *J. Virol.* **76**, 8276–8284 (2002).
- Yu, X.G. et al. Consistent patterns in the development and immunodominance of human immunodeficiency virus type 1 (HIV-1)-specific CD8⁺ T-cell responses following acute HIV-1 infection. *J. Virol.* **76**, 8690–8701 (2002).
- Novitsky, V. et al. Magnitude and frequency of cytotoxic T-lymphocyte responses: identification of immunodominant regions of human immunodeficiency virus type 1 subtype C. *J. Virol.* **76**, 10155–10168 (2002).
- Addo, M.M. et al. Comprehensive epitope analysis of human immunodeficiency virus type 1 (HIV-1)-specific T-cell responses directed against the entire expressed HIV-1 genome demonstrate broadly directed responses, but no correlation to viral load. *J. Virol.* **77**, 2081–2092 (2003).
- Altfield, M.E. Influence of HLA-B*57 on clinical presentation and viral control during acute HIV-1 infection. *AIDS* **17**, 2581–2591 (2003).
- Yang, Z. Maximum likelihood estimation based on large phylogenies and analysis of adaptive evolution in human influenza virus A. *J. Mol. Evol.* **51**, 423–432 (2000).
- McMichael, A. & Kieneman, P. HIV/AIDS. HLA leaves its footprint on HIV. *Science* **296**, 1410–1411 (2002).
- Frahm, N. et al. Identification of highly immunodominant regions in HIV by comprehensive CTL screening of ethnically diverse populations. *J. Virol.* (in press).
- Barber, L.D. et al. Polymorphism in the α 1 helix of the HLA-B heavy chain can have an overriding influence on peptide-binding specificity. *J. Immunol.* **158**, 1660–1669 (1997).
- Kelleher, A.D. et al. Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B*27-restricted cytotoxic T lymphocyte responses. *J. Exp. Med.* **193**, 375–386 (2001).
- Richman, D.D., Wrin, T., Little, S.J. & Petropoulos, C.J. Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proc. Natl. Acad. Sci. USA* **100**, 1144–1149 (2003).
- Cohen, J. HIV Escape artist par excellence. *Science* **299**, 1505–1508 (2003).
- Migueles, S.A. et al. The differential ability of HLA B*5701:1 long-term nonprogressors and progressors to restrict human immunodeficiency virus replication is not caused by loss of recognition of autologous viral gag sequences. *J. Virol.* **77**, 6899–6908 (2003).
- Novitsky, V. et al. HIV evolution dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science* **267**, 483–489 (1995).
- Greene, W.C. & Peterlin, B.M. Charming HIV's remarkable voyage through the cell: basic science as a passport to future therapy. *Nat. Med.* **8**, 673–680 (2002).
- Novitsky, V. et al. Human immunodeficiency virus type 1 subtype C molecular phylogeny: consensus sequence for an AIDS vaccine design? *J. Virol.* **76**, 5435–5451 (2002).
- Singh, A.R., Hill, R.L. & Lingappa, J.R. Effect of mutations in Gag on assembly of immature human immunodeficiency virus type 1 capsids in a cell-free system. *Virology* **279**, 257–270 (2001).
- Furuta, R. et al. HIV-1 capsid mutants inhibit the replication of wild-type virus at both early and late infection phases. *FEBS Lett.* **415**, 231–234 (1997).
- Kragstad, P. et al. Human immunodeficiency virus nucleocapsid protein polymorphisms modulate the infectivity of RNA packaging mutants. *Virology* **294**, 282–288 (2002).
- Forsey, B.M., von Schweder, U.J., Sundquist, W.I. & Aiken, C. Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. *J. Virol.* **76**, 5667–5677 (2002).
- Bratton, D., Franke, E.K. & Luban, J. Cyclophilin A is required for the replication of group M human immunodeficiency virus type 1 (HIV-1) and simian immunodeficiency virus SIVCPZ/GAB but not group O HIV-1 or other primate immunodeficiency viruses. *J. Virol.* **70**, 4220–4227 (1996).
- Bratton, D., Franke, E.K. & Luban, J. Cyclophilin A is required for an early step in the cycle of human immunodeficiency virus type 1 before the initiation of reverse transcription. *J. Virol.* **70**, 3551–3560 (1996).
- Yoo, S. et al. Molecular recognition in the HIV-1 capsid/cyclophilin A complex. *J. Mol. Biol.* **289**, 780–795 (1997).
- Goulder, P.J. et al. Substantial differences in specificity of HIV-specific cytotoxic T cells in acute and chronic HIV infection. *J. Exp. Med.* **193**, 161–194 (2001).
- Bunce, M. et al. Phototyping: comprehensive DNA typing for HLA-A, B, C, DRB1, DRB3, DRB4, DRB5 & DQB1 by PCR with 144 primer mixes utilizing sequence-specific primers (PCR-SSP). *Tissue Antigens* **46**, 355–367 (1995).
- Goulder, P.J. et al. Rapid definition of five novel HLA-A*3002-restricted human immunodeficiency virus-specific cytotoxic T-lymphocyte epitopes by Elspot and sequence cloning. *In vitro* staining assays. *J. Virol.* **75**, 1339–1347 (2001).
- Ross, H.A. & Rodrigo, A.G. Immune-mediated positive selection drives human immunodeficiency virus type 1 molecular variation and predicts disease duration. *J. Virol.* **76**, 11715–11720 (2002).
- Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556 (1997).
- Swofford, D.L. in *Phylogenetic Analysis Using Parsimony* (and other methods) 4th edn. (Sinauer Associates, Sunderland, Massachusetts, 2002).
- Hayman, A. et al. Phylogenetic analysis of multiple heterosexual transmission events involving subtype B of HIV type 1. *AIDS Res. Hum. Retroviruses* **17**, 689–695 (2001).
- Karper, P. et al. The genetic diversification of the HIV type 1 gag p17 gene in patients infected from a common source. *AIDS Res. Hum. Retroviruses* **11**, 1197–1201 (1995).